LABORATOR #8

EX#0 (Un „pic” de teorie)

#1 Media unei variabile aleatoare discrete cu valori numere naturale:

$$X = \begin{pmatrix} 0 & 1 & 2 & \cdots & k & \cdots \\ p_0 & p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix};$$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\,\mathbb{P}(X = k) = \sum_{k=0}^{\infty} k\,p_k.$$

#2 Varianța unei variabile aleatoare discrete cu valori naturale:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \sum_{k=0}^{\infty} (k - \mathbb{E}[X])^2\,\mathbb{P}(X = k).$$

#3 Media unei variabile aleatoare continue de densitate $f : \mathbb{R} \to [0, \infty)$:

$$\mathbb{P}(X \in (a, b)) = \mathbb{P}(X \in [a, b]) = \int_a^b f(x)\mathrm{d}x;$$

$$E[X] = \int_{\mathbb{R}} x\,f(x)\mathrm{d}x$$

#4 Varianța unei variabile aleatoare continue de densitate $f : \mathbb{R} \to [0, \infty)$:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x)\mathrm{d}x.$$

#5 Deviația standard: $\sigma(X) \coloneqq \sqrt{\mathrm{Var}(X)}$

#6 Media produsului a două variabile aleatoare independente X și Y cu medie finită:

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

Demonstrație. Vom demonstra egalitatea pentru două variabile discrete, cu valori naturale:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{k=0}^{\infty} k\,\mathbb{P}(XY=k) \\
&= \sum_{k=0}^{\infty} k \sum_{\substack{i,j\in\mathbb{N}\\ ij=k}} \mathbb{P}(X=i\cap Y=j) \\
&= \sum_{k=0}^{\infty} k \sum_{\substack{i,j\in\mathbb{N}\\ ij=k}} \mathbb{P}(X=i)\mathbb{P}(Y=j) \quad \text{(din independență)} \\
&= \sum_{k=0}^{\infty} ij \sum_{\substack{i,j\in\mathbb{N}\\ ij=k}} \mathbb{P}(X=i)\mathbb{P}(Y=j) \\
&= \sum_{i=0}^{\infty} i\,\mathbb{P}(X=i) \sum_{j=0}^{\infty} j\,\mathbb{P}(Y=j) = \mathbb{E}[X]\,\mathbb{E}[Y].
\end{aligned}$$

□

#7 *Legea numerelor mari:* Fie $X_1, X_2, \ldots, X_N$ variabile aleatoare independente și având aceeași distribuție (i.e., $\mathbb{P}(X_i \leq b)) = \mathbb{P}(X_j \leq b)$, pentru orice $b \in \mathbb{R}$ și orice $i, j \in \overline{1, N}$). Presupunem că avem $\mathbb{E}[X_1] < \infty$, iar, din faptul că $X_1, X_2, \ldots, X_N$ sunt identic distribuite, rezultă $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \ldots = \mathbb{E}[X_N]$.

Atunci, dacă notăm:

$$\bar{X}_N := \frac{X_1 + X_2 + \cdots + X_N}{N},$$

are loc:

$$\mathbb{P}\left(\lim_{N\to\infty} \bar{X}_N = \mathbb{E}[X_1]\right) = 1.$$

Cu alte cuvinte,

$$\bar{X}_N \to \mathbb{E}[X_1] \quad \text{aproape sigur.}$$

Mai mult, dacă $\mathrm{Var}(X_1) < \infty$, atunci pentru orice $\varepsilon > 0$ are loc estimarea:

$$\mathbb{P}\left(\left|\bar{X}_N - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X_1)}{N\varepsilon^2}.$$

#8 *Aproximarea numerică a varianței (Varianța Eșantionului):* Fie X_1, X_2, ..., X_N variabile aleatoare independente și având aceeași distribuție, cu $\mu = \mathbb{E}[X_1] < \infty$ și $\mathrm{Var}(X_1) < \infty$. Dacă notăm:

$$\mathrm{Var}_N := \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X}_N)^2,$$

atunci

$$\mathrm{Var}_N \xrightarrow{\text{aproape sigur}} \mathrm{Var}(X_1)$$

și

$$\mathbb{E}[\mathrm{Var}_N] = \frac{N-1}{N}\mathrm{Var}(X_1).$$

De aceea, de multe ori se preferă ca, în simulări numerice, să aproximăm varianța cu valoarea empirică:

$$\frac{N}{N-1}\mathrm{Var}_N = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X}_N)^2.$$

Demonstraţie.

$$\begin{aligned}
\mathrm{Var}_N &= \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu + \mu - \bar{X}_N)^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2 + 2(\mu - \bar{X}_N)\underbrace{\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)}_{=\bar{X}_n-\mu} + (\mu - \bar{X}_N)^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2 - (\mu - \bar{X}_N)^2 \xrightarrow{a.s.} \mathbb{E}[(X_1-\mu)^2] + 0 = \mathrm{Var}(X_1),
\end{aligned}$$

din Legea Numerelor Mari. Pentru a calcula $\mathbb{E}[\mathrm{Var}_N]$, scriem:

$$\begin{aligned}
\mathbb{E}[\mathrm{Var}_N] &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2\right] - \mathbb{E}\left[(\mu-\bar{X}_N)^2\right] \\
&= \mathrm{Var}(X_1) - \mathbb{E}\left[(\mu-\bar{X}_N)^2\right].
\end{aligned}$$

Folosim formula mediei produsului a două variabile independente pentru a scrie:

$$\begin{aligned}
\mathbb{E}\left[(\mu-\bar{X}_N)^2\right] &= \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^{N}(X_i-\mu)\right)^2\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\sum_{i=1}^{N}(X_i-\mu)^2\right] + \frac{2}{N^2}\mathbb{E}\left[\sum_{1\le i<j\le N}(X_i-\mu)(X_j-\mu)\right] \\
&= \frac{1}{N}\mathrm{Var}(X_1) - \frac{2}{N^2}\sum_{1\le i<j\le N}\underbrace{\mathbb{E}[X_i-\mu]}_{=0}\,\underbrace{\mathbb{E}(X_j-\mu)]}_{=0} \\
&= \frac{1}{N}\mathrm{Var}(X_1).
\end{aligned}$$

$\square$

#9 *Teorema Limită Centrală:* Fie X_1, X_2, ..., X_N variabile aleatoare independente și având aceeași distribuție, cu $\mu = \mathbb{E}[X_1] < \infty$ și $\sigma \coloneqq \sqrt{\mathrm{Var}(X_1)} < \infty$. Atunci:

$$\frac{\sqrt{N}}{\sigma}(\bar{X}_N - \mu) \xrightarrow{\text{în distribuție}} \mathcal{N}(0,1),$$

ceea ce înseamnă că, pentru orice $b \in \mathbb{R}$,

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{\sqrt{N}}{\sigma}(\bar{X}_N-\mu) \le b\right) = \mathbb{P}\left(\mathcal{N}(0,1)\le b\right).$$

EX#1 (Medie, varianță și deviație standard)

#1 Pentru fiecare dintre distribuțiile Bernoulli(p), Binomială(m, p), Geometrică(p), Poisson(λ), Uniformă($[a, b]$), Exponențială(λ), calculați teoretic media și varianța.

#2 Calculați empiric media și varianța conform Legii Numerelor Mari, respectiv a formulei Varianței Eșantionului. Afișați într-un sistem de axe de coordonate xOy eroarea dintre valoarea teoretică și aproximarea numerică, în funcție de numărul de simulări $N \in \{1, 2, \ldots, 100000\}$. Considerați, spre exemplu, $p = 0.3$, $n = 10$, $\lambda = 20$.

#3 Pentru fiecare distribuție enumerată mai sus, generați $K = 10000$ de seturi de simulări independente $X_1, X_2, \ldots X_N$, cu $N = 50000$, și generați histograma expresiei:

$$\frac{\sqrt{N}}{\sigma}\left(\frac{X_1 + X_2 + \cdots + X_N}{N} - \mu\right),$$

unde μ și σ reprezintă media, respectiv deviația standard a acelei distribuții. Suprapuneți peste histogramă graficul funcției de densitate a distribuției $\mathcal{N}(0, 1)$,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$